Transgenic Research (2005) 14:627–643
DOI 10.1007/s11248-005-5695-5

Optimization of *Acidothermus cellulolyticus* endoglucanase (E1) production in transgenic tobacco plants by transcriptional, post-transcription and post-translational modification

Ziyu Dai[1,*], Brian S. Hooker[1], Ryan D. Quesenberry[1] & Steven R. Thomas[2]
[1]*Chemical and Biological Processing Development Group, Process Science and Engineering Division, Pacific Northwest National Laboratory, Richland, WA, 99352, USA*
[2]*National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO, 80401, USA*

Received 15 December 2004; accepted 12 April 2005

Key words: Acidothermus cellulolyticus, cellulases, E1 endoglucanase, post-transcriptional modification, sub-cellular compartmentation, transgenic tobacco

Abstract

An attempt was made to obtain a high-level production of intact *Acidothermus cellulolyticus* endoglucanase (E1) in transgenic tobacco plants. The E1 expression was examined under the control of the constitutive and strong Mac promoter or light-inducible tomato Rubisco small sub-unit (*RbcS-3C*) promoter with its original or Alfalfa Mosaic Virus (AMV) *RNA4* 5′-untranslated leader (UTL) and targeted to different sub-cellular compartments via transit peptides. The transit peptides included native E1, endoplasmic reticulum, vacuole, apoplast, and chloroplast. E1 expression and its stability in transgenic plants were determined via E1 activity, protein immunoblotting, and RNA gel-blotting analyses. Effects of sub-cellular compartments on E1 production and its stability were determined in transgenic tobacco plants carrying one of six transgene expression vectors, where the *E1* was under the control of Mac promoter, mannopine synthase transcription terminator, and one of the five transit peptides. Transgenic tobacco plants with an apoplastic transit peptide had the highest average E1 activity and protein accumulation, which was about 0.25% of total leaf soluble proteins estimated via E1 specific activity and protein gel blots. Intercellular fluid analyses confirmed that E1 signal peptide functioned properly in tobacco cells to secret E1 protein into the apoplast. By replacing *RbcS-3C* UTL with AMV *RNA4* UTL E1 production was enhanced more than twofold, while it was less effective than the mannopine synthase UTL. It was observed that *RbcS-3C* promoter was more favorable for *E1* expression in transgenic plants than the Mac promoter. E1 activity in dried tobacco seeds stored one year at room temperature was 45% higher than that observed immediately after harvesting, suggesting that E1 protein can be stored at room temperature for a long period. E1 stability in different sub-cellular compartments and the optimal combination of promoter, 5′-UTL, and sub-cellular compartmentation for heterologous protein production in transgenic plants are discussed.

Introduction

The ability to use plant biomass as feedstock for conversion to other useful chemicals requires a complete hydrolysis of plant cell wall polysaccharides to fermentable sugars. Cellulose is an unbranched homopolymer of β-1, 4-linked glucose and it is the major polysaccharide component in plant biomass. Complete hydrolysis of cellulose requires at least three different hydrolytic enzymes: β-1, 4-endoglucanase (or endo-β-1, 4-glucanohydrolase; E.C. 3.2.1.4), β-1, 4-exoglucanase (or

*Author for correspondence
E-mail: ziyu.dai@pnl.gov

cellobiohydrolase; E.C. 3.2.1.91), and β-D-glucosidase (E.C. 3.2.1.21). Recycling of the carbon in plant cell walls via the action of these enzymes serves an essential function in the global carbon cycle.

Characterization, production and application of microbial cellulase systems have been studied extensively (Tucker et al., 1989; Baker et al., 1994; von Ossowski et al., 1997; Divne et al., 1998; Dai et al., 2000a,b; Ziegler et al., 2001; Ziegelhoffer et al., 2001). According to the current technology, cellulase cost is about $0.30 to $0.50 per gallon of ethanol produced, which accounts for more than 50% of the ethanol production cost and is prohibitively expensive for cost-effective ethanol production (Biomass plan, 2004). Effective and inexpensive means of cellulase production on a large scale reasons the use of cellulases as one of the key components to convert low value lignocellulosic biomass to ethanol and other chemicals, which is an environmentally benign process. In an effort to demonstrate technical feasibility and potential to reduce the cost of production of cell wall degrading enzymes, alternative production of these products in transgenic plants has been examined recently (Herbers et al., 1995; Jensen et al., 1996; Liu et al., 1997; Dai et al., 1999,2000a,b; Ziegler et al., 2001; Ziegelhoffer et al., 2001). The ultimate objective of these studies is to produce high levels of active, recoverable, and intact cellulases and the accumulation of which does not compromise plant growth and development.

The expression of a particular heterologous gene and subsequent production of its protein in plant cells are influenced by various factors. These include (1) transcriptional factors such as transgene copy number, site of integration of the T-DNA in the chromosome, and promoter activity; (2) post-transcriptional factors including mRNA splicing, 5′-untranslated leader (UTL), 3′-end formation, mRNA stability, and translation; (3) post-translational factors, for example protein stability, modification, and trafficking (Koziel et al., 1996; Gallie, 1998). Numerous studies have demonstrated that high expression levels of heterologous genes can be achieved in plants via transcriptional, post-transcriptional, and/or post-translational modifications (Jobling and Gehrke, 1987; Fiedler et al., 1997; Dai et al., 2000b; Ziegler et al., 2001; Ziegelhoffer et al., 2001). Promoters and enhancers (*cis*-acting regulatory elements) also play a crucial role in control of production of heterologous proteins at a particular plant growth and development stage or condition, or in a specific plant tissue (de Jaeger et al., 2002). For example, the tomato Rubisco small sub-unit promoter (*RbcS-3C*) confers photosynthetic tissue specificity and light regulation (Sugita et al., 1987; Gittins et al., 2000; Dai et al., 2000a). Recently, Outchkourov et al. (2003) has compared in detail the effects of different Rubisco small sub-unit (*RbcS*) promoters isolated from various plants on protein expression in different host plants to the commonly used Cauliflower Mosaic Virus (CaMV) 35S promoter. They found that all *RbcS* promoters examined yielded much higher levels of protein expression than that of 35S promoter.

Transgene expression of a thermostable endoglucanase (E1) gene, cloned from *Acidothermus cellulolyticus*, has been previously examined in transgenic *Arabidopsis,* potato, and tobacco plants (Dai et al., 2000a,b; Ziegler et al., 2001; Ziegelhoffer et al., 2001). In this study, we examined *E1* expression in transgenic tobacco plants under the control of the constitutive, strong Mac promoter (Comai et al., 1990) or the light inducible tomato Rubisco small sub-unit (*RbcS-3C*) promoter (Sugita et al., 1987) in combination with various 5′-UTL and transit peptide sequences. This has allowed a detailed comparison of the effects of different transit peptides, 5′-UTL, and promoters on E1 protein production and its stability in transgenic tobacco plants.

Experimental procedures

Bacterial strains, plant material, plant transformation, and plant growth conditions

Escherichia coli strains MC1000 and JM83 (*ara, leu, lac, gal, str*) were used as the hosts for routine cloning experiments. *Agrobacterium tumefaciens* LBA-4404 containing the Ach5 chromosomal background and a disarmed helper-Ti plasmid pAL-4404 (Hoekema et al., 1983) was used for transformation of tobacco plants (*Nicotiana tabacum* L. cv. Petit Havana SR1). Transgenic plants were obtained by the co-cultivation method (An et al., 1988) using tobacco leaf discs grown aseptically on Murashige and Skoog agar (MS) medium supplemented with 3% sucrose,

appropriate levels of plant growth regulators (Murashige and Skoog, 1962) and 50 mg/l kanamycin. The seeds of T0 transgenic plants were germinated on MS medium agar containing 50 mg/l kanamycin, and the healthy kanamycin-resistant T1 plants were picked and grown in the growth room. Growth and maintenance of transgenic tobacco plants in the plant growth room mainly followed the procedures described by Dai et al. (2000a).

Recombinant DNA techniques

Standard procedures were used for recombinant DNA manipulation (Sambrook et al., 1989). Plasmid pMPT4-5 containing a genomic clone of the *E1* gene isolated from *A. cellulolyticus* genomic library (GenBank Accession No. U33212) was obtained from the National Energy Renewable Laboratory in Golden, Colorado. The 1566 bp fragment (containing the mature peptide coding region) was isolated from pPMT4-5 by high-fidelity PCR using the primer pair P1/P2 (Table 1). Restriction endoglucanase sites Xba I and *Nsi*I were introduced in front of the start codon ATG and the *Sac*I site right after the stop codon. This *E1* PCR fragment was introduced into the transgene expression vectors Mm, Mm-vac, Mm-apo, Mm-chl, Ma-chl, Ra-chl, and Rr-chl (Figure 1). The *Xba*I site at the end of Mac or *RbcS*-3C promoter and *Sac*I site in front of transcriptional terminator Tmas or T7-T5 were, from the binary vector pZD243 derived from pAN628 (An et al., 1988). Similarly, a DNA fragment containing the *Xba*I site in front of ATG and the *Sac*I site right after the stop codon in which the entire *E1* coding region (including the signal peptide and mature polypeptide of E1 protein) was isolated by high-fidelity PCR using the PCR primers P2/P3 and introduced into the transgene expression vector Mm-E1. The DNA fragment used for transgene expression vector Mm-er, which contains KDEL right in front of the stop codon, was isolated via high-fidelity PCR with a pair of Primers P2/P4. The coding sequences for vacuole (P5/P6) and apoplast (P7/P8) transit peptides were directly synthesized, while the coding sequence for chloroplast transit peptide was isolated via high-fidelity PCR with a

Table 1. PCR primers

Construct	Primer	Oligo sequence
Mm	P1	5′-tct aga tgc atg cgg gcg gcg gct att ggc aca-3′
	P2	5′-ctt aga tct gag ctc tta act tgc tgc gca ggc gac tgt-3′
Mm–E1	P3	5′-gtc tag aat gcc gcg cgc att gcg gcg agt-3′
Mm-er	P4	5′-ctt aga tct gag ctc tta gag ttc gtc ctt act tgc tgc gca ggc gac tgt-3′
Mm-vac	P5	5′-cta gaa cca tgg ccc att cca ggt tca atc cca tcc gcc tcc cca cca cac acg aac ccg cct cct ctg aaa cta tgc a-3′
	P6	5′-tag ttt cag agg agg cgg gtt cgt gtg tgg tgg gga ggc gga tgg gat tga acc tgg aat ggg cca tgg tt-3′
Mm-apo	P7	5′-gat ccc ccg gga tga act tcc tca aaa gct tcc cct ttt atg cct tcc ttt gtt ttg gcc aat act ttg tag ctg tta ctc atg ctc tgc a-3′
	P8	5′-gag cat gag taa cag cta caa agt att ggc caa aac aaa gga agg cat aaa agg gga agc ttt tga gga agt tca tcc cgg gg-3′
Chl transit peptide	P9	5′-gg gat atc atg gct tcc tct gtc att tct tca g-3′
	P10	5′-ggg ata tcc tgc atg cag cta act ctt cca ccc tcg tca gac aaa tca gga agg-3′
Ma-chl	P11	5′-gcc gat atc tct cta tct cta cga tct agg aag-3′
	P12	5′-gaa agc ttc ccc atg gtg gaa gta ttt gaa aga aaa tta aaa ata aaa acc tgc agg gta att tgg gaa gat ata ata gga agc-3′
Ra-chl	P13	5′-ccc ctc gag gtc gac ggt atc gat atc ctt-3′
	P14	5′-gaa agc ttc ccc atg gtg gaa gta ttt gaa aga aaa tta aaa ata aaa acc tgc agg tta cta aga ggt tat tag gac gcc-3′
Rr-chl	P15	5′-gcc tct ata tgc tga aat aat tgg tta cta aga gg-3′

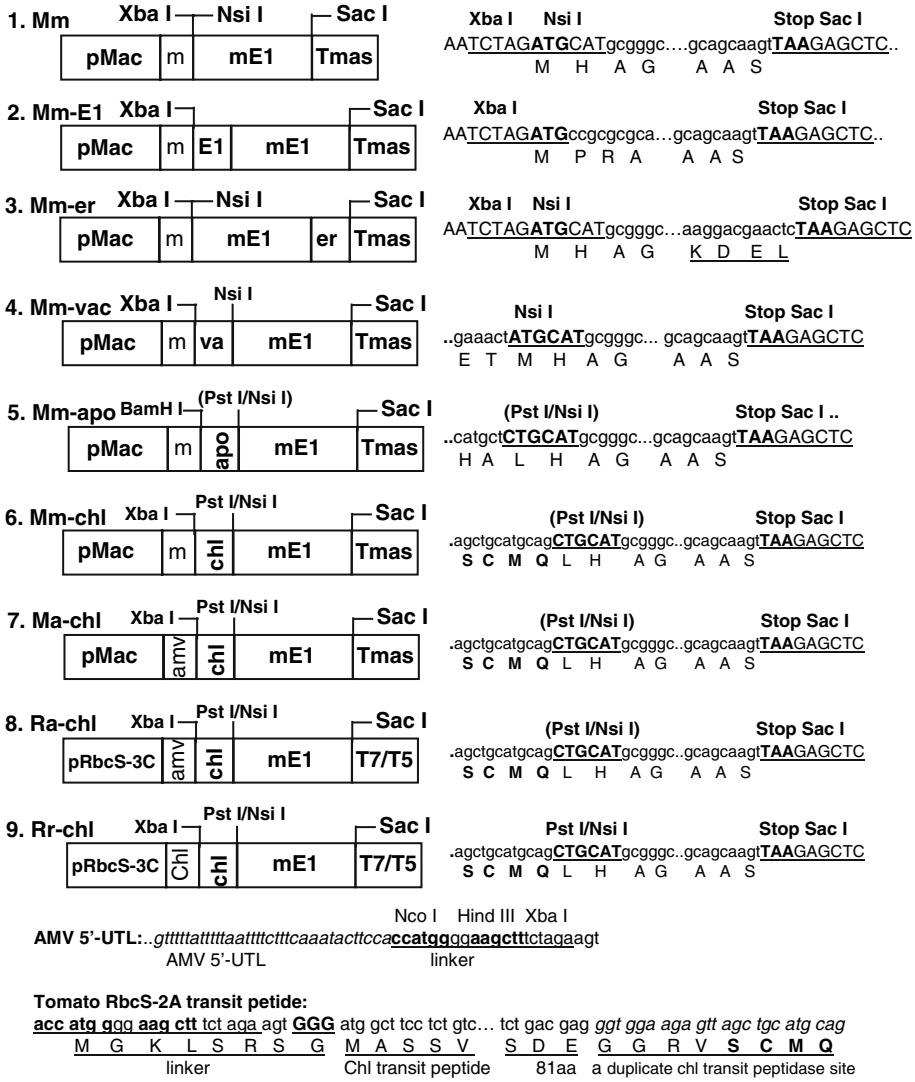

Figure 1. Schematic representation of the E1 transgene expression fragments in the binary vectors used for *Agrobacterium*-mediated transformation. Chimeric *E1* genes are the result of linking either the tomato Rubisco small sub-unit (*RbcS*-3C) or the Mac promoter, to a variety of 5′-UTL (mas [m], AMV [a], or RbcS-3C [r]), transit peptides (native E1 [E1], KDEL for the endoplasmic reticulum [er]; sweet potato sporamin A for the vacuole [vac], tobacco PR-S for the apoplast [apo], or tomato Rubisco small sub-unit 2A for the chloroplast [chl]), the mature *E1* coding sequence, and one of two transcription terminators (mas or T7-T5). Restriction endonuclease *Nsi*I, *Pst*I, *Sac*I, and *Xba*I sites were incorporated into the DNA sequence via PCR. The following nine transgene expression vectors were constructed: Mm, Mm-E1, Mm-er, Mm-vac, Mm-apo, Mm-chl, Ma-chl, Ra-chl, and Rr-chl. The detailed nucleotide and amino acids are shown around the translation start and stop sites, the AMV *RNA4* 5′-UTL and chl transit peptide with a duplicate peptidase site.

pair of primers P9/P10 using tomato *RbcS*-2A cDNA as DNA template, which consists of the first 81 amino acids of *RbcS*-2A protein followed by eight amino acids with six amino acids before the transit peptidase cleavage site. The 5′-UTL sequences of Mac and *RbcS*-3C promoters in transgene expression vectors (Ma-chl and Ra-chl) were replaced by Alfalfa Mosaic Virus (AMV) *RNA4* UTL via PCR with primers P11/P12 and P13/P14, respectively, where the first seven nucleotides (AAATTAC) of mas 5′-UTL were fused with the AMV 5′-UTL, while the whole 5′-UTL of *RbcS*-3C was replaced by AMV 5′-UTL via PCR. The *RbcS*-3C promoter in transgene expression vector Rr-chl was isolated via high-fidelity PCR with a pair of primers P13/P15.

Five different transit peptides examined in this study were native E1, endoplasmic reticulum (KDEL), vacuole (sporamin A), apoplast (pathogenic resistance gene S, PR-S), and chloroplast transit peptide (*RbcS*-2A). It has been demonstrated that the PR-S signal peptide reliably targets heterologous proteins to the apoplast (Cornelissen et al., 1986; Verwoerd et al., 1995). The sporamin A transit peptide has been well studied (Manzara et al., 1991; Matsuoka and Nakamura, 1991; Matsuoka et al., 1995) and shown to target heterologous proteins to the vacuole. It has also been well documented that the carboxy-terminal endoplasmic reticulum (ER) retention signal, KDEL, functions as an ER retention signal for foreign proteins (Chrispeels, 1991; Schouten et al., 1996; Fiedler et al., 1997). The chloroplast *RbcS*-2A signal peptide (Pichersky et al., 1986; Sugita et al., 1987) was also utilized to target proteins to the chloroplast (Dai et al., 2000a).

The expression of mature E1 gene was placed under the control of the hybrid Mac promoter (Comai et al., 1990) or the tomato *RbcS*-3C promoter (Sugita et al., 1987). In two of the chimeric E1 expression constructs (Ma-chl and Ra-chl), a synthetic 5′-UTL sequence derived from AMV *RNA4* replaced the original UTL of mannopine synthase gene and *RbcS*-3C. All PCR products were confirmed by DNA sequencing analysis. All chimeric E1 genes were terminated by the transcriptional termination sequences of either the mannopine synthase gene or the T7 and T5 genes of the octopine type Ti plasmid from *A. tumefaciens*. The binary vectors were mobilized into *A. tumefaciens* LBA 4404 by the freeze-thaw method (Holsters et al., 1978) and further utilized in *Agrobacterium*-mediated leaf disc transformation.

RNA preparation and gel blotting analysis

Total RNA was obtained from 0.1 to 1.0 g of leaf tissue essentially according to the manufacturer's instructions for the RNeasy kit (Qiagen, Inc., Valencia, CA). Total RNA concentration was quantified spectrophotometrically. About 20 μg of each total RNA sample was used for gel blot analysis as described by Dai and An (1995).

Enzyme extraction, assays, SDS-PAGE, and immunoblotting

The third or fourth leaf from the shoot apex was used for protein extraction. Leaf samples were harvested at 2–3 h into the light period. Leaf tissues were cut into approximately 1 cm^2 pieces and pooled for homogenization. The enzyme extract, and assay, SDS-PAGE, and western blot were described previously (Dai et al., 2000a).

Isolation of E1 protein from the intercellular fluid and E1 activity assays

The third or fourth leaf counted from plant apex of transgenic plants, which carried the Ra-chl, Mm-E1 and Mm-apo transgene, was harvested. One half of the leaf tissues were sliced into 1 cm × 2 cm pieces and the other half was used for direct extraction as described above. About 0.15 g of leaf pieces was vacuum-infiltrated with 50 mM MES (pH 5.5) twice each for 10 min at 20 in. of mercury. The infiltrated leaf pieces were transferred into 1.5 ml microcentrifuge tubes and centrifuged at 350 g for 10 min to obtain fluid from the intercellular space. About 15–25 μl of intercellular fluid were used for E1 activity measurement and 30–50 μl of intercellular fluid was used for protein quantification using the methods described above.

Results

A series of E1 gene expression vectors were constructed using selected combinations of two promoters (*RbcS*-3C and Mac, a hybrid promoter of mannopine synthase promoter and CaMV 35S promoter enhancer region, designated "R" and "M", respectively), three 5′ UTL [UTL: mas (mannopine synthase gene), AMV (Alfalfa Mosaic Virus *RNA4*), and *RbcS*-3C, designated "m", "a", and "r", respectively], five transit peptides (native E1, KDEL, *RbcS*-2A, PR-S, and Sporamin A designated "E1", "er", "chl", "apo", and "vac", respectively), and two transcriptional terminator sequences [Tmas (mannopine synthase gene) and T7-T5 (T7/T5 *Agrobacterium* gene)] (Figure 1). The first six E1 protein expression vectors (designated Mm, Mm-E1, Mm-er, Mm-vac, Mm-apo, and Mm-chl) varied only by the transit peptide

sequences employed and were used to determine effects of differential transit peptides on E1 protein production and stability in transgenic tobacco. In these six constructs the mature E1 coding sequence was under the control of the Mac promoter, mas UTL, the mas transcriptional terminator and one of five different transit peptides or without any transit peptide where the transgene expression vector Mm did not possess a transit peptide and allowed the E1 protein to accumulate in the cytosol.

The effect of the promoter and the UTL on *E1* gene expression were also compared using the vector 6 (Mm-chl), 7 (Ma-chl), 8 (Ra-chl) and 9 (Rr-chl), where the mature E1 coding sequence was controlled by the Mac or *RbcS-3C* promoter, its original or the AMV *RNA4* UTL, the *RbcS*-2A chloroplast transit peptide, and the Tmas or T7-T5 transcription terminator.

The effect of transit peptide on E1 production and stability in T0 transgenic tobacco plants

Proteolytic degradation of heterologous proteins is one of the main limiting factors to achieving high levels of foreign protein accumulation in non-seed tissues of plants. In order to determine effects of subcellular compartmentation on E1 protein production, the E1 protein was targeted to five different organelles via transit peptides, where the coding sequence of transit peptide and mature E1 protein was under the control of Mac promoter and mannopine synthase gene transcription terminator. A series of tobacco transgenic plants (designated Mm, Mm-E1, Mm-er, Mm-vac, Mm-apo, and Mm-chl) were generated. The E1 enzyme activity of leaf extracts of different tobacco transformants was measured as shown in Figure 2. About 20–30 transgenic plants were analyzed for each transgene expression vector. For comparison, the highest E1 activities from 15 transgenic plants per transgene expression vector were chosen. As expected, every transgenic plant exhibited a unique E1 expression level. However, comparison of the activity ranges provides a good indication of the effects of different transit peptide sequence and their sub-cellular localization on E1 protein production. Among the transgenic plants carrying one of these six transgene expression vectors, the transgenic plants containing the mature *E1* gene with the apoplast transit peptide sequence (Mm-apo) had the highest E1 activity. The E1 activity in more than 50% of Mm-apo transformant leaf extracts was over 4000 pmol 4-methylumbelliferyone (MU) mg^{-1} total leaf soluble protein min^{-1}. The highest E1 activity in the Mm-apo transformant family was 18,056 pmol MU mg^{-1} total leaf soluble protein min^{-1}, with an overall average activity of 6903 pmol MU mg^{-1} total leaf soluble protein min^{-1} in 34 transgenic plants examined (inset, Figure 2).

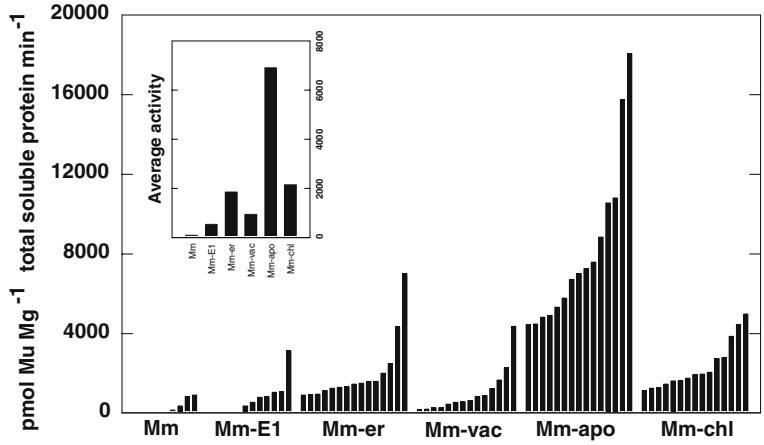

Figure 2. E1 activity of different individual T1 transgenic plants carrying one of six transgene expression vectors (Mm, Mm-E1, Mm-er, Mm-vac, Mm-apo, and Mm-chl). The mature E1 coding sequence was under the control of the Mac promoter, mannopine synthase gene 5′-UTL, mannopine synthase gene transcription terminator, and one of the following transit peptide sequences: no transit peptide (Mm), native E1 (Mm-E1), ER retention signal KDEL (Mm-er), vacuole (Mm-vac), apoplast (Mm-apo), and chloroplast (mm-chl). The inserted bar graph was the overall average of E1 activity of transgenic plants carrying one of six transgene expression vectors. MU, Methylumbelliferone.

Reasonably high E1 activity levels were observed in both Mm-er and Mm-chl transgenic plants. More than 45% of Mm-chl transgenic plants had E1 activity in the leaf extracts ranging from 1100 to 5000 pmol MU mg^{-1} total leaf soluble protein min^{-1}, with an overall average activity of 2152 pmol MU mg^{-1} total leaf soluble protein min^{-1}. Similarly, more than 36% of Mm-er transgenic plants had E1 activity ranging from 1100 to 7000 pmol MU mg^{-1} total leaf soluble protein min^{-1}, with an overall average activity of 1853 pmol MU mg^{-1} total leaf soluble protein min^{-1}. However, E1 activity in leaf extracts from most Mm-E1 and Mm-vac transgenic plants was less than 2000 pmol MU mg^{-1} total leaf soluble protein min^{-1}, with an overall average activity of 553 and 954 pmol MU mg^{-1} total leaf soluble protein min^{-1}, respectively. The Mm transgenic plants had the lowest E1 activity in the leaf extracts among the six different transgenic groups. Of 25 Mm transgenic plants tested, more than 50% of them had no detectable E1 activity.

The effect of the AMV RNA4 5′ untranslated leaders and promoter on E1 expression in T0 transgenic tobacco

It has been reported that mRNA stability and translation can be enhanced by 5′-UTL such as AMV *RNA4* and Tobacco Mosaic Virus in transgenic plants (Gallie et al., 1987; Jobling and Gehrke, 1987). The effect of AMV *RNA4* UTL on the E1 gene expression was compared to that of the tomato *RbcS*-3C gene (Sugita et al., 1987) or of the mannopine synthase gene (Barker et al., 1983). The E1 activity was measured in leaf extracts from about 20 transgenic plants harboring one of transgene expression frames Mm-chl, Ma-chl, Ra-chl or Rr-chl. In these four transgene expression frames, the mature E1 coding sequence was under the control of Mac or *RbcS-3C* promoter, either the mannopine synthase gene 5′-UTL (Mm-chl), *RbcS*-3C 5′-UTL sequence (Rr-chl) or AMV *RNA4* 5′-UTL sequence (Ma-chl and Ra-chl), the *RbcS*-2A transit peptide sequence and the mas or T7-T5 transcription terminator. The Ra-chl transgenic plants had the highest E1 activity in leaf extracts compared to the Mm-chl, Ma-chl, and Rr-chl transgenic plants examined (Figure 3). The highest E1 activity in Ra-chl transgenic plants was 15,622 pmol MU mg^{-1} total leaf soluble protein

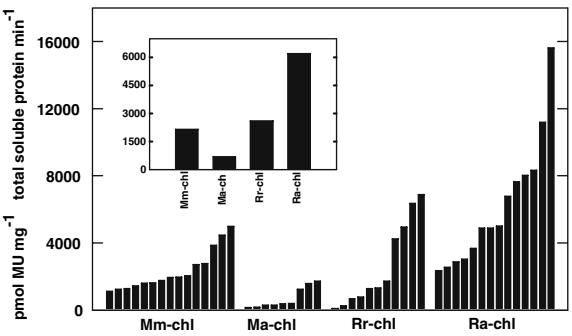

Figure 3. E1 activity of different individual transgenic plants carrying different transgene expression vectors (Mm-chl, Ma-chl, Rr-chl, and Ra-chl). The mature E1 coding sequence was under the control of the Mac, or tomato Rubisco small subunit (RbcS-3C) gene promoters, with respective native 5′-UTLs or the Alfalfa Mosaic Virus RNA 4 5′-UTL (a), and chloroplast transit peptide (chl) sequence, generating Mm-chl, Ma-chl, Rr-chl, and R-chl transgene expression vectors.

min^{-1} and overall average activity was 6204 pmol MU mg^{-1} total leaf soluble protein min^{-1} over 22 transgenic plants examined. Among the 16 Rr-chl transgenic plants examined the highest E1 activity was 6886 pmol MU mg^{-1} total leaf soluble protein min^{-1} and overall average activity 2594 pmol MU mg^{-1} total leaf soluble protein min^{-1}. The results show that the AMV *RNA4* UTL indeed increased the E1 production in Ra-chl transgenic plants as compared with that of Rr-chl transgenic plants, which is consistent with previous observations (Gallie et al., 1987; Jobling and Gehrke, 1987).

The averaged E1 activity in the leaf extracts of Mm-chl transgenic plants was 2153 pmol MU mg^{-1} total leaf soluble protein min^{-1} and the highest E1 activity was 3122 pmol MU mg^{-1} total leaf soluble protein min^{-1}. However, the E1 activity in the leaf extracts of Ma-chl transgenic plants, in which the 5′-UTL of manopine synthase was replaced by the AMV *RNA4* UTL, was the lowest one among these four sets of transgenic plants, with an average E1 activity of 692 pmol MU mg^{-1} total leaf soluble protein min^{-1}.

Analysis of T1 transgenic plants carrying different transgene expression vectors with E1 activity, protein immunoblot, and RNA blot analyses

E1 activity. Since the transgenic plants carrying one of the nine transgene expression vectors were generated at different times, the age and

physiological state of the resulting transgenic plants may have varied from batch to batch. Variations in the growth conditions may have also been one of the factors for experimental variation. In order to compare *E1* expression among these transgenic lines with similar growth conditions, kanamycin-resistant, uniform and healthy T1 transgenic plants were selected and grown from seeds harvested from self-pollinated T0 plants having the highest *E1* expression in each of the nine different transformant families. The E1 activity in leaf extracts of T1 transgenic plants was measured. Eight kanamycin resistant sibling plants were assayed for each set of T1 progeny (Figure 4). In each transformant family, a range of expression levels was found among the siblings due to genetic segregation. The range of expression levels in segregating T1 families reflected the presence of E1 hetero- and homozygotic individuals. In the Mm-er, Mm-vac, Mm-chl, Ra-chl, and Rr-chl families, one or two with higher expression are clearly differentiated from the rest of their siblings and could represent the homozygotic individuals. In the other families, two classes of expression patterns were not distinct suggesting that the E1 activity varies among individuals of the same genotype and that the activity in some heterozygotes was close to that of the homozygotes. Those differences were more likely to be caused by individual variations of E1 expression than by errors in our measurements since all plants grew under the same conditions and E1 activity analysis was completed for all plants at the same time. The highest averaged E1 activities in T1 transgenic families was the Ra-chl and was followed by Mm-apo, Rr-chl, Mm-chl, Mm-er, Mm-E1, Mm-vac, Mm-chl, and Mm, respectively.

E1 immunoblots and E1 protein stability in different organelles. In order to establish if E1 activity correlates with E1 protein accumulation and to determine E1 stability at different subcellular compartments, 2–5 individual T1 transgenic plants with relatively higher E1 activity from each transgenic family were selected for protein immunoblotting. The E1 activity for selected transgenic plants was determined first (shown in Figure 5, upper panel) and the same protein extracts were used for protein immunoblotting analysis. Corresponding protein immunoblots with the monoclonal antibody raised against *A. cellulolyticus* E1 protein are shown in Figure 5 (lower panel). The intensity of immunoblots in Figure 5 clearly correlated with the E1 activities in the leaf tissues examined. The Ra-chl transgenic family, which had the highest E1 activity, possessed the highest level of E1 protein accumulation in leaf tissues, while the tranformant Mm family possessing the lowest E1 activity in leaf extracts also had the lowest level of E1 protein accumulation. However, the accumulation of E1 protein in Ra-chl transgenic plants was about 10 higher than that of Mm-apo transgenic based on the protein

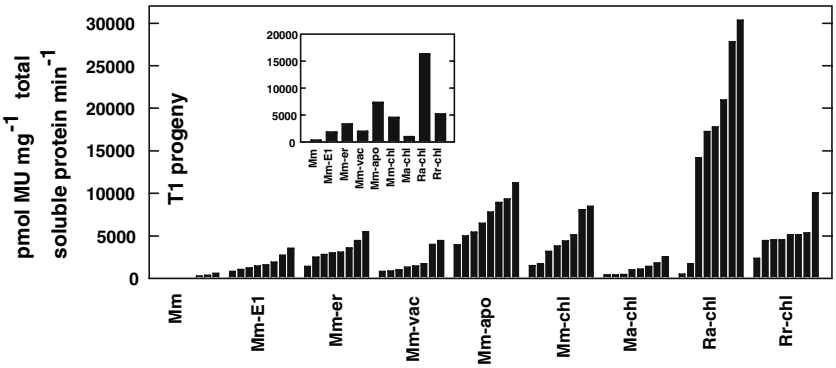

Figure 4. E1 activity of individual transgenic plants carrying different transgene expression vectors from T1 progeny. The E1 coding sequence was under the control of the Mac promoter and mas UTL, without any transit peptide sequence (Mm), native E1 signal peptide sequence (Mm-E1), ER retention signal peptide (KDEL) sequence (Mm-er), vacuolar transit peptide sequence (Mm-vac), apoplast signal sequence (Mm-apo), and chloroplast transit peptide sequence (Mm-chl). The E1 coding sequence was also under the control of the Mac promoter and AMV UTL with a chloroplast transit peptide sequence (Ma-chl) or RbcS-3C promoter, AMV *RNA4* UTL, and chloroplast transit peptide (Ra-chl) or RbcS-3C promoter, its original UTL, and the chloroplast transit peptide sequence (Rr-chl).

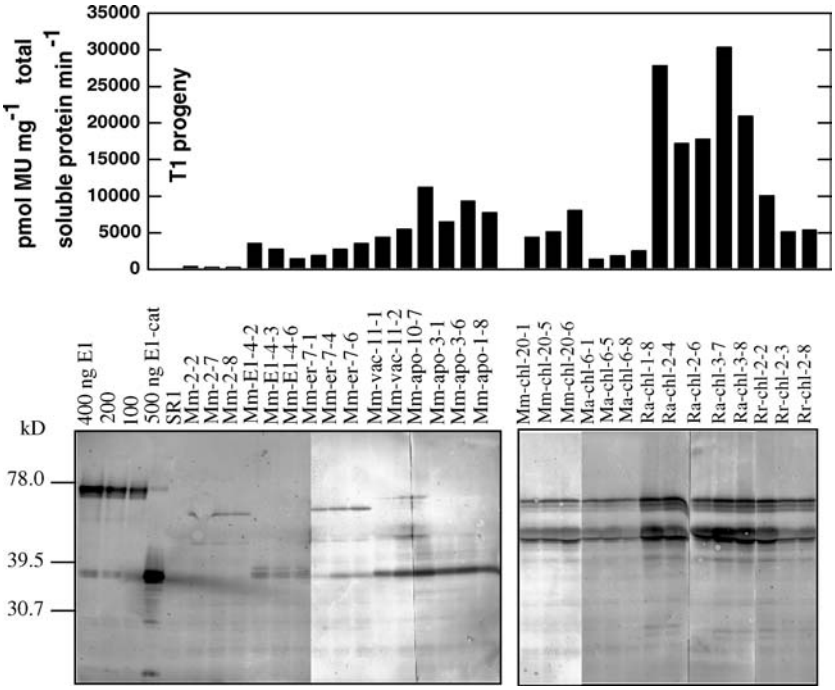

Figure 5. Western blot analysis for E1 protein expressed in leaf tissues of selected transgenic plants. The upper panel shows the E1 activity of selected transgenic plants used for western blot analysis in the lower panel. The lower panel is a western blot using 30 μg of total soluble protein from different leaf extracts that were separated via PAGE, transferred onto 0.45 μm nitrocellulose membrane and immunoblotted. The first four lanes show the full-length E1 protein (lanes 1–3) or E1 catalytic domain (lane 4) isolated from bacterial culture. Lane 5 represents the tobacco control plant and lanes 6–8 E1 protein isolated from Mm transgenic lines. Lanes 9–11 represent the E1 protein isolated from Mm-E1 transgenic lines and lanes 12–14 the E1 protein isolated from Mm-er transgenic lines. Lanes 15 and 16 represent the E1 protein isolated from Mm-vac transgenic lines and lanes 17–20 E1 protein isolated from Mm-apo transgenic lines. Lanes 21–23 represent the E1 protein isolated from Mm-chl transgenic lines and lanes 24–26 was the E1 protein isolated from Ma-chl transgenic lines. Lanes 27–31 represent the E1 protein isolated from Ra-chl transgenic lines and lanes 3234 the E1 protein isolated from mm-apo transgenic lines.

gel blot intensity, while the E1 activity in Ra-chl transgenic plants was only about two times higher than that in Mm-apo transgenic plants. The reason for this discrepancy is not known.

In addition, E1 protein immunoblots in Figure 5 (lower panel) also show distinct stepwise degradation patterns of E1 protein in the transgenic tobacco plants carrying different transgene expression vectors with specific transit peptides. The immunoblots of E1 protein from leaf extracts of Mm transgenic plants (Lanes 6–8), carrying the transgene expression frame Mm that contains no signal peptide, shows a distinct single band which is smaller than that of E1 protein (Lanes 1–3) purified from a *Streptomyces lividans* (a Gram-positive soil bacterium) expression system containing a signal peptide (42 aa) and mature E1 protein (521 aa) (Figure 5, lower panel). The E1 protein immunoblot for Mm-E1 transgenic plants (Lanes 9–11) shows a distinct stepwise degradation pattern down to about 38 kD, which is of similar size as the E1 catalytic domain. In the Mm-er transgenic plants (Lanes 12–14), E1 protein immunoblots showed two clear bands in which one was slightly larger than that in the Mm transgenic plants and the other was of similar size to the E1 catalytic domain (Lane 4). The E1 protein in Mm-vac (Lanes 15 and 16) and Mm-apo (Lanes 17–20) transgenic plants showed similar degradation pattern, in which all E1 proteins were degraded to a molecular weight of about 38 kD, similar to the E1 catalytic domain. The E1 protein in Mm-chl Lanes 21–23), Ma-chl (Lanes 24–26), Ra-chl (Lanes 27–31), and Rr-chl (Lanes 32–34) transgenic plants possessed similar degradation patterns (Figure 5, lower panel) since those four transgene expression vectors contain the same 89 amino acids of RbcS-2A transit peptide consisting of 57 amino acids of its transit peptides,

the first 24 amino acids of the mature protein and the eight amino acids of transit peptidase cleavage site (Figure 1).

The stepwise degradation pattern of the E1 protein in the Rr-chl transgenic plants, representing the four chloroplast targeting (-chl) transgene expression vectors, was compared to the E1 protein purified from *S. lividans* and Mm and Mm-E1 transgenic plants (Figure 6). The first two bands of the E1 protein isolated from the Rr-chl transgenic plants are larger than the E1 protein purified from Mm transgenic plants and the third band was similar to it. Another three distinct bands larger than the E1 catalytic domain also exist, which is larger than the E1 protein isolated from Mm-E1 transgenic plants. All six stepwise degradation bands are smaller than that from *S. lividans*.

RNA blots. The transcription level of the E1 gene in selected transgenic plants was also determined using RNA gel blots. Results in Figure 7 show that E1 coding sequence, which was under the control of different promoters, UTLs, transit peptide sequences, and transcriptional terminators in the nine transgenic families, was transcribed to the predicted range of 1.56–1.85 kb mRNA. The Ra-chl transgenic plants, which have the highest E1 activity and protein accumulation in leaf tissues, have the highest E1 transcription level compared to other transgenic families.

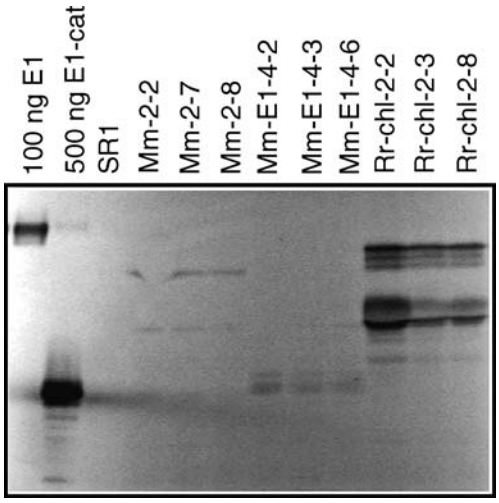

Figure 6. E1 protein stepwise degradation between transgenic plants Mm, Mm-E1, and Ra-chl. Lane 1 is the E1 protein purified from bacterial culture and Lane 2 is the E1 catalytic domain expressed in bacteria culture. Lane 3 is the protein extracts from SR1 tobacco plant. Lanes 4–6 show E1 proteins isolated from the leaves of Mm transgenic plants. Lanes 7–9 show the E1 proteins isolated from Mm-E1 transgenic plant leaf tissues. Lanes 10–12 show E1 protein isolated from leaf tissues of Ra-chl transgenic plants. About 30 μg of total soluble protein was separated by PAGE in each lane.

Confirmation of E1 protein targeting via native E1 transit peptide and PR apoplast transit peptide in selected transgenic plants

The results in Figure 5 show that the stepwise degradation pattern of E1 proteins in Mm-E1 and Mm-apo transgenic plants were quite similar. In order to determine whether E1 protein was properly localized to the apoplast (intercellular space), which contain the native E1 or PR apoplast transit peptide, the E1 activity in the intercellular fluid isolated from one half of the leaf tissues and in the ground leaf extract obtained from other half of leaf tissues was determined and compared based on the method described by Verwoerd et al. (1995). The Ra-chl and Rr-chl transgenic plants were selected as a control, where E1 was targeted into the chloroplasts (Dai et al., 2000b). The E1 activity in the intercellular fluid and the total leaf extracts for Ra-chl, Rr-chl, Mm-E1, and Mm-apo was determined and the ratio of E1 activity in intercellular fluid and total leaf extract was calculated and shown in Figure 8. In the control of Ra-chl and Rr-chl transgenic plants, the ratio of E1 activity in the intercellular fluid over the total leaf extract was only around 1.7–2.4. In the Mm-E1 transgenic plants the ratio was 8–24, while it was 70–170 for Mm-apo transgenic plants.

E1 protein stability in dried tobacco seeds

E1 protein stability in dried leaves of the Ra-chl transgenic tobacco plants has been examined previously, where E1 protein was partially degraded at the first 15 h of dehydration and remained similar for the rest of 100 h of treatment (Dai et al., 2000b). In this study, the E1 activity in the newly harvested and 12-month old stored seeds of transgenic plants carrying the transgene Mm-apo was determined. The results in Figure 9 show that after 12 months of storage, the E1 activity was 6727 pmol MU mg^{-1} total soluble protein min^{-1}, which is about 48% higher than that immediately after harvested dried seeds (Figure 9).

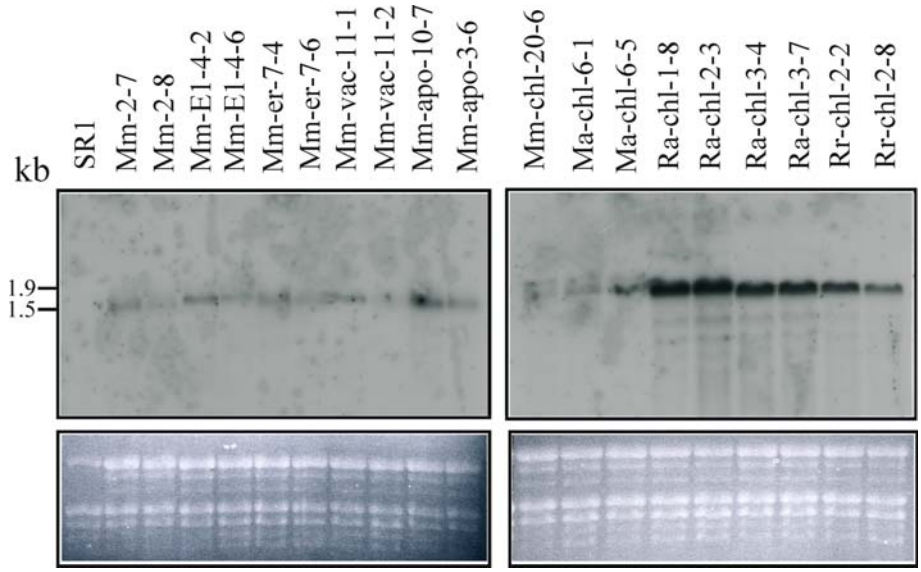

Figure 7. The RNA gel-blotting to determine E1 transcription levels in select transgenic tobacco plants with high E1 activity from different transgenic lines. The RNA gel-blots contain 20 μg per lane of total RNA isolated from selected T1 progeny transgenic plants probed using 1.2 kb Xba I/BamH I E1 coding sequence fragments labeled with [α-^{32}P]-dCTP. The RNA blot was developed from X-ray film exposed in a −30°C freezer for 40 h.

Discussion

Producing a high yield of recombinant, intact, and fully functional protein is the ultimate goal in many applied plant biotechnology programs. It is also important to synthesize the recombinant protein in such a way that it may be stably stored for a relatively long time in plant tissues. However, the level of protein expression and protein stability in plants can be influenced by many factors that range from transgene copy number, chromosomal location, *cis*-regulatory elements, mRNA stability, modification of protein product, protein trafficking, and final compartmentalization in plant tissues (Koziel et al., 1996; Conrad and Fiedler, 1998; Gallie, 1998; Dai et al., 2000a; Ziegelhoffer et al., 2001; Outchkourov et al., 2003). Therefore, a generally adopted strategy to increase heterologous protein accumulation levels in plants is to employ a proper combination of promoter, 5′-end UTL and transcriptional terminator and target the expressed protein for retention in the endoplasmic reticulum or other organelles of proper tissues. In different sub-cellular compartments of non-seed tissues, a wide variety of proteases are involved in modification and degradation of proteins. Levels of these proteases are affected by many biotic and abiotic factors such as the growth and development of vegetative and reproductive tissues.

The cost-effective production of cellulase enzymes to catalyze the hydrolysis of cellulose to sugars is still one of the hierarchical barriers to fully realizing enzyme-based sugar production technology, one of the important aspects in establishing biomass as a significant source of sustainable fuels, heat, power, chemicals, and other materials. Several labs have attempted to improve the expression of cellulase enzymes in both microbial (Park et al., 1998; Pauly et al., 1999; Bergquist et al., 2004) and higher plant systems (Herbers et al., 1995; Dai et al., 1999, 2000a,b; Ziegler et al., 2001; Ziegelhoffer et al., 2001; Bailey et al., 2004). Previously, we examined E1 expression in tobacco and potato under the control of the Mac or RbcS-3C promoter combined with native E1, apoplastic, chloroplastic, or vacuolar targeting and a mannopine synthase (Mas) or T7/T5 transcriptional terminator (Dai et al., 2000a,b). In those earlier studies, E1 protein accumulation in tobacco and potato leaf tissues ranged up to 1.35–2.6% total soluble protein in leaf extracts. Stepwise degradation pattern of E1 protein, observed in transgenic potato plants with E1 protein targeted to chloroplasts, was

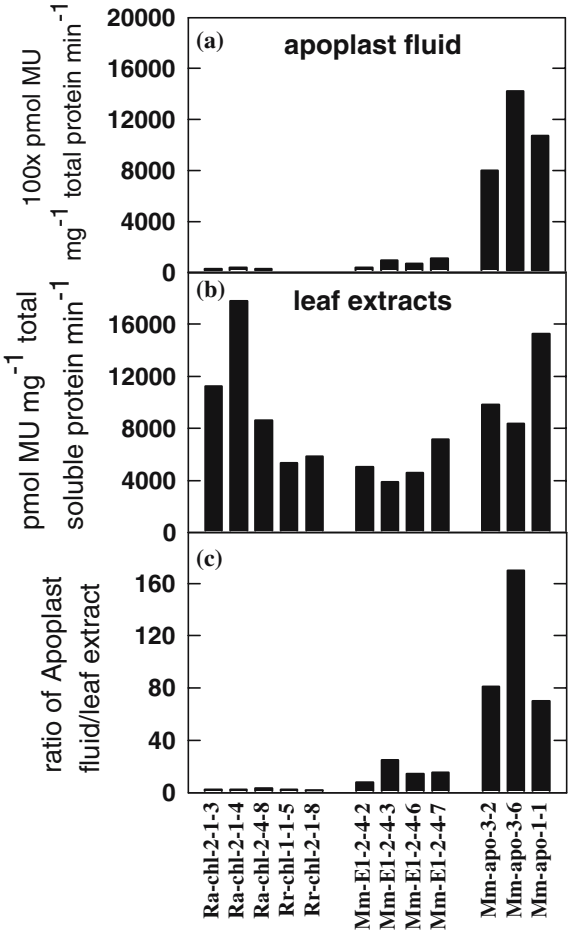

Figure 8. The E1 activity in the apoplastic fluid and leaf extracts of select T1 progeny transgenic plants carrying Ra-chl, Mm-E1, or Mm-apo transgene. The E1 activity was determined: (a), in the apoplast fluid isolated from one half of leaf segments by vacuum infiltration in 50 mM MES pH5.5 buffer; (b), in the leaf extracts of the remaining the other half of leaf tissues by grinding in liquid N_2 and E1 protein extraction buffer; (c), the ratio of E1 activity in apoplastic fluid of leaf tissues versus in total leaf extract. Selected T1 transgenic plant Ra-chl-2-1-3, Ra-chl-2-1-4, 2-4-8, Ra-chl-1-1-5, and Ra-chl-2-1-8 was fron Ra-chl transgene. Selected T1 transgenic plant Mm-E1-2-4-2, Mm-E1-2-4-3, Mm-E1-2-4-6, and Mm-E1-2-4-7 was fron Mm-E1 transgene. Selected T1 transgenic plant Mm-apo-3-2, Mm-apo-3-6, and Mm-apo-1-1 was fron Mm-apo transgene.

completely different from that with E1 protein targeted to apoplasts. The E1 protein targeted to the chloroplast was degraded to several protein fragments with the size larger than 45 kD, while with targeting to the apoplast it was completely degraded to about 39.5 kD protein fragment. The E1 protein targeted to the vacuole was barely detectable in transgenic potato. Ziegler et al.

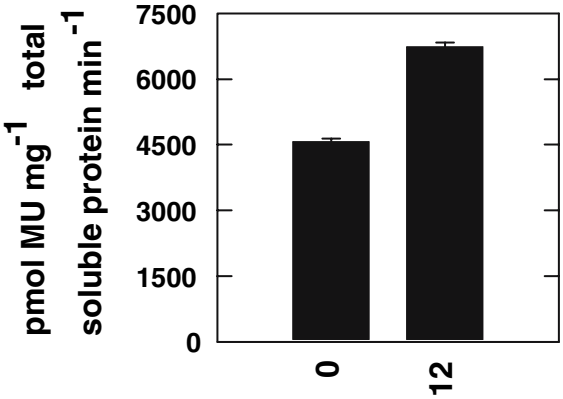

Figure 9. The E1 activity in total soluble protein extracts isolated from dried seeds stored for 0–12 months at room temperature and low humidity. Total soluble protein in dried seeds was isolated by grinding in the liquid N_2–frozen mortar and pestle and extracting in E1 extraction buffer.

(2001) examined expression of E1 catalytic domain that was targeted to the apoplast in *Arabidopsis thaliana* and observed high level of E1 accumulation in the apoplasts. Ziegelhoffer et al. (2001) evaluated the expression of the E1 full-length gene versus the catalytic domain in transgenic tobacco, which was under the control of CaMV 35S promoter, apoplast or chloroplast transit peptide sequence, and nopaline synthase gene transcription terminator. The E1 protein accumulation in the cytosol, chloroplast, or apoplast was partially or completely degraded to the fragments with the size similar to E1 catalytic domain, which confirmed the observation in the mm-apo transgenic potato plants (Dai et al., 2000b).

In this study, effects of promoter and the 5′-UTL on E1 gene expression and effects of different sub-cellular compartmentalization on E1 protein production and stability in transgenic tobacco were compared in significant detail. To analyze the optimal location for high-level accumulation and the stability of E1 protein in specific sub-cellular compartments, we transformed tobacco plants using six different transgene constructs, in which E1 gene was under the control of Mac promoter, mas UTL, mas transcriptional terminator, and one of five transit peptide sequences. Fusion with a transit peptide is sufficient for entry of protein into the secretory system by translocation across the ER membrane. Further transport depends on a specific targeting or retention signal and transport competence (Chrispeels, 1991; Vitale

and Denecke, 1999). Proteins that lack targeting or retention information are secreted via the bulk flow pathway into the cytosol (Denecke et al., 1990; Chrispeels, 1991; Vitale and Denecke, 1999). It has been demonstrated that the pathogenesis related gene S (PR-S) transit peptide (Verwoerd et al., 1995), chloroplast (RbcS-2A) transit peptide (Dai et al., 2000b), sporamin gene signal peptide (Matsuoka and Nakamura, 1991; Matsuoka et al., 1995), and KDEL (Conrad and Fiedler, 1998) are able to target or retain heterologous proteins to apoplast, chloroplast, vacuole, and ER, respectively.

When we analyzed E1 activity in leaf protein extracts of Mm, Mm-E1, Mm-er, Mm-vac, Mm-apo, and Mm-chl transgenic plants (Figures 2, 4, and 5), the Mm-apo transgenic plants had the highest E1 activity with an overall average of about 7295 pmol MU mg^{-1} total soluble protein min^{-1} in the T1 progeny, which was 28, 4, 4, 2.2, and 1.6 times higher than that in Mm, Mm-E1, Mm-vac, Mm-er, and Mm-chl transgenic plants, respectively (Figure 4), while the mRNA levels in those transgenic plants were similar (Figure 7), suggesting that E1 protein production was mainly affected by sub-cellular compartmentation. Among these six transgene expression, Mm transgenic plants, where the signal (transit) peptide of the E1 gene was removed and the coding region carried no transit peptide, accumulated the lowest amount of E1 protein in the cytosol (Figure 5, Lanes 6–8 of left panel) with two visible polypeptide fragments of E1 protein whose molecular weight were smaller than that purified from *S. lividans*. This suggests that there exists a set of proteases in the cytosol that can rapidly degrade full-length E1 protein and that the localization of E1 protein to the cytosol should be avoided. The E1 protein in all other transgenic plants was partially or completely degraded to different polypeptide fragments. The E1 protein in transgenic plants carrying transgene expression frames Mm-E1, Mm-vac and Mm-apo, which contain native E1 protein, vacuole (sporamin A), and apoplast (PR-S) transit peptides, respectively, had a consistent protein degradation pattern, in which all E1 protein was degraded to a polypeptide slightly larger than the E1 catalytic domain. This may suggest that E1 protein was degraded during the sorting process or that the vacuole and apoplast possess similar proteases.

The proper function of native E1 protein signal peptide and PR transit peptide was evaluated via intercellular fluid analysis. The Ra-chl and Rr-chl transgenic plants known for chloroplast targeting (Dai et al., 2000a) were used as control. Estimating the ratio of E1 activity in intercellular fluid over the whole leaf tissues of those transgenic plants shows that Ra-chl and Rr-chl plants had only about 2, while was 8–24 and 70–170 in the Mm-E1 and Mm-apo transgenic plants, respectively. It suggests that the transit peptides of E1 protein and PR protein indeed lead the E1 protein secretion into the apoplast. However, the secretion of E1 protein into the apoplast via the PR transit peptide was much more effective than that via the E1 protein transit peptide since the ratio of E1 activity in the Mm-apo transgenic plants is about seven times higher than that of the Mm-E1 transgenic plants. This difference may be due to the origin of the PR and E1 protein. The results also suggest that the bacterial protein transit peptide can function properly in higher plants.

In the Mm-er transgenic plants part of the E1 protein was degraded to the fragment, which was similar to that in Mm-E1, Mm-vac and Mm-apo transgenic plants. This degradation may be due to the existence of proteases in the ER or leakage in which the E1 protein might be secreted to apoplast or vacuole compartments where it was degraded. Indeed, it has been observed that the C-terminal KDEL sequence of the KDEL-tailed cysteine proteinase may be involved in the formation of KDEL vesicle for vacuolar transport (Okamoto et al., 2003).

Previously, Dai et al. (2000a) observed microscopically that E1 protein, which was fused to tomato RbcS-2A transit peptide (57 amino acids) and the first 24 amino acids of mature RbcS-2A polypeptide with an addition of eight amino acids of transit peptidase cleavage site, was properly sorted into the chloroplast. In this study, E1 protein stepwise degradation patterns of transgenic tobacco plants with chloroplast targeting was compared to other transgenic plants with different targeting signals (Figures 5 and 6). The first two bands are larger than the mature E1 protein in the Mm transgenic plants where there was no transit peptide in E1 protein, while the third band is similar in size with the E1 protein in Mm transgenic plants. The fourth and fifth bands are smaller than the E1 protein in the Mm

transgenic plants, but much larger than the E1 catalytic domain. The first largest band might be the E1 protein fused with the first 89 amino acids of *RbcS*-2A consisting of 57 amino acids of *RbcS*-2A transit peptide, the first 24 amino acids of its mature protein and the eight amino acid of repeated transit peptidase site. The second band might be due to the partial removal of *RbcS*-2A transit peptide. When the E1 protein was sorted into the chloroplast it was further partially degraded to the fourth and fifth bands due to the existence of certain proteases inside the chloroplasts. Further examining those protease sites within the E1 protein via protein sequencing will allow us to further improve E1 protein production in transgenic plants.

The effect of the AMV *RNA4* 5′-UTL on E1 transgene expression was examined using both Mac and *RbcS*-3C promoters. The UTL sequence of eukaryotic mRNA plays a major role in translation efficiency (Koziel et al., 1996). It has been demonstrated that AMV coat protein 5′-UTL enhances mRNA translation (Jobling and Gehrke, 1987; Browning et al., 1988). The effect of a 5′-UTL may vary depending on the plant, particularly between dicots and monocots (Gallie and Young, 1994; Koziel et al., 1996). In this study we found that an overall average of E1 activity in the Ra-chl transgenic plants, in which the UTL of AMV *RNA4* replaced the UTL of the *RbcS*-3C, was three times higher than that in Rr-chl transgenic plants containing the original UTL of *RbcS*-3C (Figures 3 and 4). Similarly, Ra-chl transgenic plants had higher E1 gene transcription and E1 protein accumulation than Rr-chl transgenic plants (Figures 5 and 7). In contrast, when the UTL of the mannopine synthase gene in Mac promoter was replaced with the UTL of AMV *RNA4*, the E1 activity and E1 protein accumulation in the Ma-chl transgenic plants was 3–4 times lower than that in the Mm-chl transgenic plants (Figures 3–5) though the transcription levels were similar in the selected transgenic plants (Figure 7). Hann and Gehrke (1995) also observed that the UTL of AMV *RNA4* did not enhance the translation of α-globin mRNA in lysates from poliovirus-infected HeLa cells. Recently, Nagao and Obokata (2003) observed that when the (U)20 motif was inserted into the 5′-UTL of the *gusA* reporter gene, GUS activity/*gusA* mRNA ratio was 5.6-fold higher in the (T)20-inserted GUS transformants than in the controls, while the (A)20, (G)20, and (C)20 motif did not affect the GUS activity/*gusA* mRNA ratio. The underlying mechanisms for this observation are unknown. In the 35 base pairs of AMV-*RNA4* UTL, there also exists a fragment containg 17 T (U) out of 21 nucleotides at its 5′-end of the UTL region. The whole UTL in *RbcS*-3C promoter was replaced by AMV *RNA4* UTL, while the region from +8 to the end of 5′-UTL (+61) of Mac promoter was replaced with AMV *RNA4* 5′-UTL. The E1 activity/*E1* mRNA ratio in Ra-chl transgenic plants was much higher than that in Rr-chl transgenic plants, while the ratio in Ma-chl transgenic plants was much lower than that in Mm-chl transgenic plants suggesting that the insertion position of a heterologous 5′-UTL may affect protein expression efficiency.

Furthermore, the effect of Mac and tomato *RbcS*-3C promoters on E1 transgene expression was examined. The activity of Mac promoter, consisting of the mannopine synthase gene (mas) region from +65 to −301 fused to the 35S promoter upstream enhancer region from −90 to −941, was much higher than that of the double 35S promoter (Kay et al., 1987) in transgenic tobacco based on GUS activity analysis (Comai et al., 1990). The photosynthetic tissue specific promoter, *RbcS*-3C, is a very useful promoter for the control of transgene expression in green tissues. When maize sucrose-phosphate synthase was expressed in tomato under the control of the tobacco Rubisco small sub-unit or CaMV 35S promoter, the level of sucrose-phosphate synthase was enhanced as much as 80% by the 35S promoter while it was unchanged under the control of tobacco RbcS promoter (Laporte et al., 2001). In our previous study however, the E1 expression in transgenic potato under the control of tomato *RbcS*-3C promoter and UTL of AMV *RNA4* was much higher than that under the control of the Mac promoter and mannopine synthase UTL (Dai et al., 2000b). In order to define whether the enhancement of E1 expression was due to the *RbcS*-3C promoter or the UTL of the AMV *RNA4*, E1 expression in transgenic tobacco, under the control of the Mac promoter or the tomato *RbcS*-3C promoter, was compared. The averaged E1 activity in the transgenic plants under the control *RbcS*-3C (Rr-chl) promoter was slightly higher (about 1.2-fold) than that under the control

of the Mac promoter (Mm-chl) (Figures 3 and 4). The level of E1 gene transcription and protein accumulation in selected transgenic plants was also slightly higher than that of selected Mm-chl transgenic plants (Figures 5 and 7). This suggests that the promoter is only one of multiple key factors in the regulation of transgene expression.

In summary, this study illustrated the effects of different sub-cellular compartments [cytosol (Mm), E1 (Mm-E1), ER (Mm-er), vacuole (Mm-vac), apoplast (Mm-apo), and chloroplast (Mm-chl)] on E1 protein expression, its accumulation and stability under the same (Mac) promoter and transcription terminator (Tmas) and one of the transit peptides. The Mm-apo transgenic plants accumulated the highest level of E1 protein (about 0.25% of total soluble leaf proteins based on the E1 specificity activity). No phenotypic changes were observed in E1-expressing transgenic plants. Four distinctly stepwise degradation patterns of E1 protein were observed in these targeted compartments. Most E1 protein in Mm transgenic plants was larger than 50 kD and in Mm-E1, Mm-vac, and Mm-apo transgenic plants similar or slightly larger than the E1 catalytic domain. The E1 protein in Mm-er transgenic plants had two E1 fragments where the large one was similar to that in Mm transgenic plants and the small one was similar to that in the Mm-E1, Mm-vac, and Mm-apo transgenic plants. There were at least five distinct E1 fragments in the transgenic plants with chloroplast targeting. Three of them were larger than or similar to that in Mm transgenic plants and two were much larger than the E1 catalytic domain. These suggest that there exist different proteases in different sub-cellular compartments and secretion pathways. The E1 secretion to the apoplast via transit peptides of E1 protein or PR-S protein was confirmed via intercellular fluid analyses. The effects of UTL of AMV *RNA4*, tomato *RbcS-3C*, and mannopine synthase gene on E1 protein expression were studied. UTL of AMV *RNA4* gene enhanced E1 protein expression when it replaced the entire UTL of *RbcS*-3C, but was less effective than the UTL of mannopine synthase gene in Mac promoter suggesting the potential of position effects of U rich region in protein translation efficiency that may be controlled via unknown underlying mechanisms. The results also showed that tomato *RbcS-3C* promoter was better than the constitutive, strong Mac promoter for *E1* expression in transgenic tobacco plants. The E1 enzyme in the seeds stored at room temperature for one year remained very active based on E1 activity analysis. Results of this study clearly indicate that E1 expression under the control of *RbcS-3C* promoter, AMV *RNA4* UTL and T7/T5 transcription terminator and with chloroplast targeting was the optimal combination for E1 protein production in tobacco plants. Further determination of E1 degraded fragments via peptide sequencing will aid to generate a new modified E1 protein with protease resistance for high production in transgenic plants.

Acknowledgements

We thank Dr. Wilhelm Gruissem, Institute of Plant Science, Swiss Federal Institute of Technology, Switzerland, for providing the two *E. coli* strains containing the *RbcS*-3C promoter (pRbcS-3C-101-1) and *RbcS*-2 cDNA [PTss-1-91 (#2)-IBI] and Dr. L. Comai, Botany Department, University of Washington, for providing the Mac promoter (pLAY112). We also thank Dr. Mike Himmel, Bill Adney and Dr. Rafael Nieves for generously providing purified E1 and E1-cat enzymes as well as the monoclonal antibody to E1. This work was conducted using funds provided by a sub-contract from the Biofuels Program of the United States Department of Energy, Office of Fuels Development. The Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RL01830.

References

An G, Ebert PR and Ha S-B (1988) Binary vectors. In: Gelvin SB and Schilperooort RA (eds), Plant Molecular Biology Mannual. (pp. 1–19) Kluwer Academic Publishers, Dordrecht, The Netherland.

Bailey MR, Woodard SL, Callaway E, Beifuss K, Magallanes-Lundback M, Lane JR, Horn ME, Mallubhotla H, Delaney DD, Ward M, GFVan, Howard JA and Hood EE (2004) Improved recovery of active recombinant laccase from maize seed. *Appl Microbiol Biotechnol* **63**: 390–397.

Baker JO, Adney WS, Nieves RA, Thomas SR, Wilson DB and Himel ME (1994) A new thermostable endoglucanase,

Acidothermus cellulolyticus E1: synergism with Trichoderma reesei CBHI and comparison to thermomonospora fusca E15. *Appl Biochem Biotechnol* **45–46**: 245–256.

Barker RF, Idler KB, Thompson DV and Kemp JD (1983) Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine Ti plasmid pTi15955. *Plant Mol Biol* **2**: 335–350.

Bergquist PL, Te'o VS, Gibbs MD, Curach NC and Nevalainen KM (2004) Recombinant enzymes from thermophilic microorganisms expressed in fungal hosts. *Biochem Soc Trans* **32**: 293–297.

Biomass plan (2004) The US Department of Energy Biomass program-multi-year technical plan.

Browning KS, Lax SR, Humphreys J, Ravel JM, Jobling SA and Gehrke L (1988) Evidence that the 5'-untranslated leader of mRNA affects the requirement for wheat germ initiation factors 4A, 4F, and 4G. *J Biol Chem* **263**: 9630–9634.

Chrispeels MJ (1991) Sorting of proteins in the secretory system. *Annu Rev Plant Physiol Plant Mol Biol* **42**: 21–53.

Comai L, Moran P and Maslyar D (1990) Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. *Plant Mol Biol* **15**: 373–381.

Conrad U and Fiedler U (1998) Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. *Plant Mol Biol* **38**: 101–109.

Cornelissen BJ, van Hooft Huijsduijnen RA and Bol JF (1986) A tobacco mosaic virus-induced tobacco protein is homologous to the sweet-tasting protein thaumatin. *Nature* **321**: 531–532.

Dai Z, Hooker BS, Anderson DB and Thomas SR (2000a) Expression of *Acidothermus cellulolyticus* endoglucanase E1 in transgenic tobacco: biochemical characteristics and physiological effects. *Transgenic Res* **9**: 43–54.

Dai Z, Hooker BS, Anderson DB and Thomas SR (2000b) Improved plant-based production of E1 endoglucanase using potato: expression optimization and tissue targeting. *Mol Breed* **6**: 277–285.

Dai Z, Hooker BS, Quesenberry RD and Gao J (1999) Expression of *Trichoderma reesei* Exo-cellobiohydrolase I in transgenic tobacco leaves and calli. *Appl Biochem Biotechnol* **77–79**: 689–699.

Dai Z and An G (1995) Induction of nopaline synthase promoter activity by H_2O_2 has no direct correlation with salicylic acid. *Plant Physiol* **109**: 1191–1197.

De Jaeger G., Scheffer S, Jacobs A, Zambre M, Zobell O, Goossens A, Depicker A and Angenon G (2002) Boosting heterologous protein production in transgenic dicotyledonous seeds using Phaseolus vulgaris regulatory sequences. *Nat Biotechnol* **20**: 1265–1268.

Denecke J, Botterman J and Deblaere R (1990) Protein secretion in plant cells can occur via a default pathway. *Plant Cell* **2**: 51–59.

Divne C, Stahlberg J, Teeri TT and Jones TA (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 A long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J Mol Biol* **275**: 309–325.

Fiedler U, Phillips J, Artsaenko O and Conrad U (1997) Optimization of scFv antibody production in transgenic plants. *Immunotechnol* **3**: 205–216.

Gallie DR, Sleat DE, Watts JW, Turner PC and Wilson TM (1987) A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression in vivo. *Nucleic Acids Res* **15**: 8693–8711.

Gallie DR (1998) Controlling gene expression in transgenics. *Curr Opin Plant Biol* **1**: 166–172.

Gallie DR and Young TE (1994) The regulation of gene expression in transformed maize aleurone and endosperm protoplasts. Analysis of promoter activity, intron enhancement, and mRNA untranslated regions on expression. *Plant Physiol* **106**: 929–939.

Gittins JR, Pellny TK, Hiles ER, Rosa C, Biricolti S and James DJ (2000) Transgene expression driven by heterologous ribulose-1, 5-bisphosphate carboxylase/oxygenase small-subunit gene promoters in the vegetative tissues of apple (*Malus pumila mill.*). *Planta* **210**: 232–240.

Hann LE and Gehrke L (1995) mRNAs containing the unstructured 5' leader sequence of alfalfa mosaic virus RNA 4 translate inefficiently in lysates from poliovirus-infected HeLa cells. *J Virol* **69**: 4986–4993.

Herbers K, Monke G, Badur R and Sonnewald U (1995) A thermostable xylanase from *Clostridium thermocellum* expressed at high levels in the apoplast of transgenic tobacco has no detrimental effects and is easily purified. *Bio/technol* **13**: 63–66.

Hoekema A, Hooykaas PJ and Schilperoort RA (1983) A binary vector strategy based on seperation vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature* **303**: 179–181.

Holsters M, de WD, Depicker A, Messens E, Van MM and Schell J (1978) ransfection and transformation of *Agrobacterium tumefaciens*. *Mol Gen Genet* **163**: 181–187.

Jensen LG, Olsen O, Kops O, Wolf N, Thomsen KK and von Wettstein D (1996) Transgenic barley expressing a protein-engineered, thermostable (1,3-1,4)-beta-glucanase during germination. *Proc Natl Acad Sci U S A* **93**: 3487–3491.

Jobling SA and Gehrke L (1987) Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. *Nature* **325**: 622–625.

Kay R, Chan A, Daly M and McPherson J (1987) Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science* **236**: 1299–1302.

Koziel MG, Carozzi NB and Desai N (1996) Optimizing expression of transgenes with an emphasis on post-transcriptional events. *Plant Mol Biol* **32**: 393–405.

Laporte MM, Galagan JA, Prasch AL, Vanderveer PJ, Hanson DT, Shewmaker CK and Sharkey TD (2001) Promoter strength and tissue specificity effects on growth of tomato plants transformed with maize sucrose-phosphate synthase. *Planta* **212**: 817–822.

Liu JH, Selinger LB, Cheng KJ, Beauchemin KA and Moloney MM (1997) Plant seed oil-bodies as an immobilization matrix for a recombinant xylanase from the rumen fungus *Neocallimastix patriciarum*. *Mol Breed* **3**: 463–470.

Manzara T, Carrasco P and Gruissem W (1991) Developmental and organ-specific changes in promoter DNA-protein interactions in the tomato rbcS gene family. *Plant Cell* **3**: 1305–1316.

Matsuoka K, Bassham DC, Raikhel NV and Nakamura K (1995) Different sensitivity to wortmannin of two vacuolar sorting signals indicates the presence of distinct sorting machineries in tobacco cells. *J Cell Biol* **130**: 1307–1318.

Matsuoka K and Nakamura K (1991) Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. *Proc Natl Acad Sci U S A* **88**: 834–838.

Murashige T and Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol Plant* **15**: 473–497.

Nagao I and Obokata J (2003) A poly(U) motif in the 5' untranslated region enhances the translational efficiency of β-glucuronidase mRNA in transgenic tobacco. *Plant Sci* **165**: 621–626.

Okamoto T, Shimada T, Hara-Nishimura I, Nishimura M and Minamikawa T (2003) C-terminal KDEL sequence of a KDEL-tailed cysteine proteinase (sulfhydryl-endopeptidase) is involved in formation of KDEL vesicle and in efficient vacuolar transport of sulfhydryl-endopeptidase. *Plant Physiol* **132**: 1892–1900.

Outchkourov NS, Rogelj B, Strukelj B and Jongsma MA (2003) Expression of sea anemone equistatin in potato. Effects of plant proteases on heterologous protein production. *Plant Physiol* **133**: 379–390.

Outchkourov NS, Peters J, de JJ, Rademakers W and Jongsma MA (2003) The promoter-terminator of chrysanthemum rbcS1 directs very high expression levels in plants. *Planta* **216**: 1003–1012.

Park JH, Oh SA, Kim YH, Woo HR and Nam HG (1998) Differential expression of senescence-associated mRNAs during leaf senescence induced by different senescence-inducing factors in Arabidopsis. *Plant Mol Biol* **37**: 445–454.

Pauly M, Andersen LN, Kauppinen S, Kofod LV, York WS, Albersheim P and Darvill A (1999) A xyloglucan-specific endo-beta-1,4-glucanase from *Aspergillus aculeatus*: expression cloning in yeast, purification and characterization of the recombinant enzyme. *Glycobiology* **9**: 93–100.

Pichersky E, Bernatzky R, Tanksley SD and Cashmore AR (1986) Evidence for selection as a mechanism in the concerted evolution of Lycopersicon esculentum (tomato) genes encoding the small sub-unit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Proc Natl Acad Sci U S A* **83**: 3880–3884.

Sambrook JF, Fritsch EF and Maniatis T (1989) Molecular Cloning: A Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NewYork.

Schouten A, Roosien J, van Engelen FA, de Jong GA, Borst-Vrenssen AW, Zilverentant JF, Bosch D, Stiekema WJ, Gommers FJ, Schots A and Bakker J (1996) The C-terminal KDEL sequence increases the expression level of a single-chain antibody designed to be targeted to both the cytosol and the secretory pathway in transgenic tobacco. *Plant Mol Biol* **30**: 781–793.

Sugita M, Manzara T, Pichersky E, Cashmore A and Gruissem W (1987) Genomic organization, sequence analysis and expression of all five genes encoding the small sub-unit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. *Mol Gen Genet* **209**: 247–256.

Tucker MP, Mohagheghi A, JGrohman K and Himmel ME (1989) Ultra-thermostable cellulases from *Acidothermus cellulolyticus*: comparison of temerature optima with previously reported cellulases. *Biotechnology* **7**: 817–820.

Verwoerd TC, Paridon PAvan, Ooyen AJvan, Lent JWvan, Hoekema A and Pen J (1995) Stable accumulation of *Aspergillus niger* phytase in transgenic tobacco leaves. *Plant Physiol* **109**: 1199–1205.

Vitale A and Denecke J (1999) The endoplasmic reticulum-gateway of the secretory pathway. *Plant Cell* **11**: 615–628.

Ossowski IVon, Terri T, Kalkkinen N and Oker-Blom C (1997) Expression of a fungal cellobiohydrolase in insect cells. *Biochem Biophys Res Com* **233**: 25–29.

Ziegelhoffer T, Raasch JA and Austin-Phillips S (2001) Dramatic effects of truncation and sub-cellular targeting on the accumulation of recombinant microbial cellulase in tobacco. *Mol Breed* **8**: 147–158.

Ziegler MT, Thomas SR and Danna KJ (2001) Accumulation of a thermostable endo-1,4-β-D-glucanase in the apoplast of Arabidopsis thaliana leaves. *Mol Breed* **6**: 37–46.